

Ein globaler „Dauerbrenner“

Der Umgang mit Hate Speech auf Online-Plattformen

Amélie Pia Heldt

2019-11-25T13:00:41

Hate Speech umgrenzen, verstehen und bekämpfen – ein weiterhin aktuelles Thema im Kontext von Online-Kommunikation. Das Thema ist demzufolge auf dem Internet Governance Forum gut aufgehoben, denn es betrifft alle globalen Kommunikationsdienste und alle Rechtsordnungen, wenn auch auf unterschiedliche Art und Weise. Doch der Umgang mit Hate Speech und die Gegenmaßnahmen zeigen bisher wenig Wirkung. Ein mehrversprechender Ansatz wäre, sich mit der Verantwortung von Plattformen zu beschäftigen, ohne sie zwangsläufig staatlichen Akteuren gleichzusetzen.

„Hate Speech“ – Definition, Interpretation, Kontext

Kaum ein Begriff aus der US-amerikanischen Rechtswissenschaft ist so präsent in der öffentlichen Debatte weltweit wie dieser [Sammelbegriff](#). Er steht für Äußerungen, die darauf abzielen, Menschen abzuwerten, anzugreifen oder anhand derer zu Hass oder Gewalt aufgerufen wird. Die Definition ist breiter als juristische Tatbestände und nicht mit ihnen gleichzusetzen: das macht den Umgang mit diesem Phänomen zur Herausforderung, denn zum einen kommt der Begriff aus einem US-Rechtssystem in dem die Meinungsfreiheit nicht durch den Gesetzgeber einschränkt werden darf (so steht es wörtlich im Ersten Zusatz zur US-Verfassung). In der Rechtsprechung des Supreme Courts sind zwar Kategorien anerkannt, die Ausnahmen zum sachlichen Schutzbereich der Meinungsfreiheit bilden, aber generell wird letzterer (für europäische Verhältnisse) sehr weit verstanden. Zum anderen variiert das Verständnis von „Hassrede“ weltweit je nach Kulturkreis und Kontext, insbesondere wenn die Kriterien nicht anhand eines objektiven Tatbestand festgelegt sind. Wenn ein Verhalten nicht verboten oder strafbar ist, kann es nichtsdestotrotz öffentlich missbilligt sein, aber die Beurteilung fällt schwerer, weil zusätzlich zu objektiven und allgemeinen Voraussetzungen subjektive Meinungen und Moralvorstellungen miteinfließen. Dieses Problem potenziert sich, wenn in einem virtuellen Raum Menschen weltweit kommunizieren können und die Plattform, die diesen Kommunikationsdienst anbietet, aus einer Rechtsordnung und einem Rechtsverständnis stammt, wo jede Regulierung von Meinungsfreiheit als Zensur betrachtet wird. In dieser Mischung von sozialen Normen und fehlenden rechtlichen Vorgaben liegt die Kernherausforderung der Content Moderation ([Roberts 2016](#)), denn soziale Netzwerke wollen bestimmen, was bei Ihnen veröffentlicht wird, aber scheuen gleichzeitig vor der Verantwortung zurück. Sie beschreiben sich selbst als Dienste, die Menschen weltweit miteinander verbinden und die Meinungsfreiheit stärken wollen, aber lehnen Inhalte wie Hassrede ab und interferieren somit auf inhaltlicher Ebene.

Nicht nur Hate Speech ist für die Online-Öffentlichkeit eine Herausforderung, sondern auch Kinderpornographie, extremistische und terroristische Inhalte, Bilder extremer Gewalt u.s.w. Gegen Kinderpornographie und [Terrorpropaganda](#) haben sich Dienstanbieter zusammengeschlossen und tauschen Informationen aus, um solche Inhalte so schnell wie möglich weltweit zu entfernen. Bei Hate Speech hingegen gibt es keinen gemeinsamen Nenner auf globaler Ebene was die genaue Definition und die Auslegung angeht – was in den USA von der Meinungsfreiheit geschützt wird, kann in Deutschland strafbar sein, wie zum Beispiel Volksverhetzung. Social Media Plattformen müssen daher eine Gradwanderung vornehmen, wenn sie ihre Dienste weltweit anbieten: sie stehen entweder als private Zensurapparate oder als Verbreiter menschenfeindlicher Inhalte da. Es ist deswegen nicht zielführend über eine globale Lösung zu sprechen, die die Inhalte anvisiert. Vielmehr muss die Staatengemeinschaft überlegen, wie viel Macht sie privaten Plattformen einräumen will, wenn die Kommunikation der Zukunft via private Infrastruktur verläuft und inhaltlich von den Dienst Anbietern bestimmt wird. Es geht um die Emergenz neuer, nicht-staatlicher Akteure, die zwar keine direkten Adressaten des Völkerrechts sind, aber für ein globales Problem verantwortlich sind, oder zumindest in hohem Maße involviert.

Löschen ja, aber filtern nein?

Lassen sich lokale Regeln weltweit anwenden, und wenn ja, wie? Das Urteil des EuGHs zur Frage, wie weit die Löschungspflicht für soziale Netzwerke bei strafbaren (nach österreichischem Recht) Inhalte reicht, ist einerseits konsequent aber andererseits bietet es keine Antwort auf die Folgeprobleme, die eine solche weltweite Anwendung beinhaltet ([Glawischnig-Piesczek/Facebook Ireland Ltd – C 18/18](#)). Das Gericht begründete seine Entscheidung damit, dass Facebook weltweit tätig ist („globale Dimension des elektronischen Geschäftsverkehrs“), die Ehrverletzung nicht nur lokal stattfindet („Schnelligkeit und geografische Ausbreitung“) und dass die Kenntnis im Sinne des [Art. 14 Abs. 1 a\) E-Commerce-Richtlinie](#) weder auf einen EU-Staat, noch auf die Mitgliedstaaten der EU begrenzt sei. Auf der Lösungsebene setzte es darauf, dass technische Lösungen vorhanden seien (Rn. 46), wobei es angesichts von [Art. 15 Abs. 1 E-Commerce](#) betonte, dass damit keine aktive Überwachung gemeint sei (Rn. 37, 42). Ja, Algorithmen können gleiche oder wesentlich gleiche Inhalte erkennen und entfernen, aber diese Systeme stoßen gerade bei Texten und der Hermeneutik an ihre Grenzen. Zum einen weil man sie durch das Hinzufügen von Wörtern „destabilisieren“ kann, zum anderen weil sie den Kontext nicht erfassen (können). Künstliche Intelligenz, d.h. zunehmend autonome Systeme, können durch maschinelles Lernen zwar immer performanter werden, aber sie stellen eine Black Box dar, deren Entscheidungsgrundlage und -vorgang nicht überprüfbar sind. Zwar wird der ursprüngliche Algorithmus von Menschen programmiert, aber zu einem gibt es noch keine [verbindlichen Standards](#) hierfür und zum anderen schwindet die Nachvollziehbarkeit je autonomer die Systeme werden.

Des Weiteren fehlt ein Hinweis des EuGH, wie mit weltweiten Löschungspflichten nur reaktiv und nicht proaktiv umzugehen sei. Der EuGH hat sich bereits vor Jahren in seinen zwei SABAM-Entscheidungen ([Scarlet/SABAM: C#70/10](#); [SABAM/Netlog](#):

C-360/10) gegen eine verdachtslose Filterung von Kommunikation ausgesprochen, doch diese Klarheit vermisst man in der jetzigen Entscheidung. Außerdem kann man bei Livestreams nicht mehr von *ex ante* oder *ex post* Löschungen sprechen: will man Bilder wie die vom [Amoklauf in Christchurch](#) (NZ) verbannen, dann muss das Hochladen von Inhalten mit einer Prüfung einhergehen, oder man verzichtet auf die simultane Übertragung. Technische Lösungen wie Upload-Filter, die die informationelle Selbstbestimmung und die Meinungsfreiheit beschränken, müssen nicht unter dem Vorwand hingenommen werden, die Datenmenge sei ansonsten nicht zu bewältigen.

Löschen ja, egal wer und wie?

Ist ein Spillover-Effekt von Gesetzen gegen Hassrede wünschenswert? Das [Netzwerkdurchsetzungsgesetz](#) (NetzDG) hat vor und nach seiner Verabschiedung (2017) für viel Diskussionsstoff und scharfe Kritik gesorgt, u.a. vom [UN-Sonderberichterstatter für Meinungsfreiheit, David Kaye](#). Das Gesetz verpflichtet soziale Netzwerke ab einer Nutzerzahl von zwei Millionen Nutzern eine Meldemöglichkeit für strafbare Inhalte in Deutschland einzurichten, solche Meldungen zu prüfen und „offensichtlich strafbare Inhalte“ innerhalb von 24 Stunden zu löschen. Ob eine strafbare Handlung vorliegt, wird demzufolge erstmal den Content Reviewers überlassen und weder die Form der Meldung, noch ein Put-Back-Recht im Falle unbegründeter Löschung wurden vorgesehen. Neben vielen anderen Kritikpunkten formeller und materieller Natur, besteht das höchste Risiko für die Meinungsfreiheit der Nutzer in dem Anreiz, Inhalte im Zweifel zu löschen um Sanktionen zu vermeiden (sog. Overblocking). Darüber hinaus wird der Umfang der zu löschenden Inhalte immer breiter, da Plattformen auf umfassende Lösungen setzen, die sie (bspw. durch Algorithmen) möglichst global einsetzen können (sog. Over-Removal). Folglich haben nationale Gesetze nicht nur im jeweiligen Land Konsequenzen, sie werden über nationale Grenzen hinaus entweder von den sozialen Netzwerken extensiv angewandt oder inklusive ihrer Mängel von anderen Staaten [übernommen](#). Dieser Übertragungseffekt ist kaum wieder einzufangen, auch nicht durch Korrekturen des deutschen Gesetzgebers. In ihrem Maßnahmenpaket [„Gegen Extremismus und Hass“](#) hat die Bundesregierung angekündigt, dass soziale Netzwerke „strafrechtlich relevante Beiträge wie Morddrohungen oder volksverhetzende Inhalte“ mitsamt der IP-Adressen der Absender dem Bundeskriminalamt melden müssen. Welche Folgen die Meldepflicht bei absoluten Antragsdelikten wie § 185 f. StGB haben soll und wer die erhöhte Zahl der [Anzeigen bewältigen](#) soll, wird daraus nicht klar. Das Vorhaben löst außerdem weder die Schwierigkeiten des NetzDGs, noch das Problem von Hate Speech.

Ausblick

Das Thema „Hate Speech“ ist von der Agenda von Internet-bezogenen Konferenzen nicht wegzudenken — leider. Auch wenn die Meinungsfreiheit aus [Art. 19 ICCPR](#) für nicht-staatliche Akteure rechtlich nicht-bindend ist und auch sonst das Verständnis von Meinungsfreiheit und ihrer Grenzen differiert, muss darauf geachtet werden, dass sie unter dem Vorwand des Schutzes der Öffentlichkeit nicht unverhältnismäßig eingeschränkt wird. Im Ergebnis wird die Antwort nicht entweder juristisch oder technologisch sein, sie wird beides beinhalten und auch Medienkompetenz und

Bildung miteinbeziehen müssen. [Ein Vorschlag](#) wäre außerdem, dass das Herstellen von Öffentlichkeit mit der Einhaltung bestimmter Grundrechte einhergehen müsse, damit wären alle Akteure – ob staatlich oder nicht – an die Mindestgewährleistung von Meinungs- und Informationsfreiheit gebunden. Eventuell werden sich soziale Netzwerke ansonsten davon verabschieden müssen, ein Spiegel der Öffentlichkeit darzustellen, wenn sie gegen Hate Speech vorgehen wollen, denn den Anspruch „allen eine Stimme zu geben“ können sie nicht erfüllen.

Amélie Pia Heldt ist Junior Researcher im Bereich Meinungsmacht & digitale Medien am Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI) in Hamburg.

Cite as: Amélie Pia Heldt, "Ein globaler 'Dauerbrenner' – Der Umgang mit Hate Speech auf Online-Plattformen", *Völkerrechtsblog*, 25. November 2019.

